The division of innovative labor among university, entrepreneurial, and corporate R&D in an analysis of agricultural biotechnology patents

Gregory D. Graff

University of California at Berkeley
and
Bio Economic Research Assoc. (bio-era™), Cambridge, MA

Theory would predict that R&D organizations in different sectors of the economy specialize—according to each one's comparative advantages—in generating innovations of different qualitative economic characteristics and at different phases in the evolution of a technology over time (i.e. in the technological trajectory). This idea of specialization by R&D sector is tested using U.S. patent data on biotechnology inventions in crop agriculture, categorized into subsets estimated to capture the most significant technological trajectories as they emerged during the formative years of the agricultural biotechnology industry. Established citations-based indices proxy for patents' heterogeneous economic and technical qualities such as value, originality, generality, and appropriability. The results of a discrete random variable regression model show systematic differences in the qualities and timing of innovations across sectors, but they are much more nuanced and interesting than a simple linear model of 'basic' public-sector and 'applied' private-sector innovation. Assuming that these innovations were, at least in some sense, optimally produced by the different sectors, these results may be interpreted as 'revealing' the underlying comparative advantages characteristic of each R&D sector.

JEL Classifications: O31, Q16, L33

The author would like to thank dissertation advisors at U.C. Berkeley, David Zilberman, Alain de Janvry, Brian Wright, and Bronwyn Hall, for their guidance and patience, as well as David Mowery and other participants in the Innovation Seminar at the Walter A. Haas School of Business, and Robert Evenson, Nicholas Kalaitzandonakis, and other commentators at the 2002 ICABR conference in Ravello, Italy, for many helpful comments and suggestions. All remaining errors and omissions are the responsibility of the author.

1. Introduction

The roles of different sectors of the R&D economy in driving innovation have long been a subject of economic study and debate. In the agricultural life sciences the question has intensified since biotechnology arose in the 1970s and 1980s out of both publicly and privately funded R&D. This new biologically based technological regime has shaken up a relatively stable *status quo* in the division of innovative labor in agriculture, challenging intersectoral relationships and giving rise to new arrangements for generating and appropriating value from innovation. Have the roles of public and private sector researchers become indistinguishable, as corporations invest in projects—such as sequencing the genomes of important crops—that can be deemed scientific public goods and while university and government laboratories make commercially valuable discoveries, such as genes, which they then privatize via patents, to be developed and marketed by private firms? Are universities and government labs being subsidized to provide substitute outputs—of knowledge and technology—in a market where industry is, in fact, *not* underinvesting? How prevalent are incumbent corporations in the creation of fundamentally new technologies and how effective are startups as a vehicle for technology development? While much has been made of these questions in policy debates and writings, little systematic empirical work has documented the qualitative differences in output between these various sources of biological innovation for agriculture.

In this chapter I suggest that one of the sources of R&D role confusion may lie in a failure to accurately reconcile evolutionary economic theories on how naturally heterogeneous technologies tend to emerge over time with the predictions of theory that organizations in different sectors of the agricultural R&D economy should enjoy relative comparative advantages at different phases in the evolution of a new technology. Existing theories on innovation suggest that basic exploratory research serves, with some probability, to create the new problem-solving paradigms that, if successful, initiate new "technological trajectories", temporal sequences of technological developments within a narrowly defined problem solving paradigm that result in new commercial process and product applications. What I am questioning, in essence, is whether the common "linear hypothesis" of innovation, when realistically cast within a probabilistic framework of heterogeneous innovation, can indeed serve to explain the different roles of public and private agricultural R&D, while still allowing for instances of public R&D yielding some private-goods-like innovations and private R&D yielding some public-goods-like innovations. A simple comparative advantage argument suggests that, as a result of different organizational endowments and characteristics of the sectors, publicly funded researchers will tend to specialize in more uncertain exploratory research and privately funded researchers will specialize in more narrowly focused, certain, and appropriable research.

I test this idea using U.S. patent data on biological inventions with relevance for crop agriculture. The database compiles information from the front pages of patents to categorize the inventions into subsets that are estimated to capture the most significant technological trajectories emerging during these formative years of the agricultural biotechnology industry. It is possible to proxy for a patent's quality, value, originality, generality, and appropriability using established citations-based indices from the NBER Patent Citations Data File (Hall, Jaffe and Trajtenberg, 2001). Assignee designations on the patents are used to identify what type of organization generated each invention—whether a government laboratory, a university, a non-profit research organization, an individual inventor, an entrepreneurial startup firm, or an established corporate firm. While both the sector of invention and the observed characteristics of a patent are endogenous within the technological trajectories framework developed here, multinomial regression allows for a partial correlation analysis capable of testing the hypotheses of sectoral specialization.

Patents are particularly useful for this exercise as they are a common measure across sectors of commercially relevant R&D output in agriculture. Whereas in other industries government and university patenting make an almost insignificant contribution—less than 3 percent on average, according to USPTO summary data—in this field of technology government and university R&D contributes upwards of 25 percent of the U.S. patents, meaning that systematic comparison across sectors is possible. The greatest drawback of using patent data is, of course, that not all inventions are patented, and differences in institutional significance of patents result in different propensities to patent across sectors. For example, economically significant inventions made at universities often show up in published research papers, not in patents, while many inventions made within companies are kept secret altogether. It is also important to note that the use of patent data necessarily constrains the investigation to issues of the original inventorship and not the current ownership of the technologies claimed in the patents. This is because a U.S. patent document lists only the name(s) of the organization(s) to whom the property rights were originally assigned when the patent was granted; neither the patent document nor the patent office keeps a running record of who currently holds title to the property rights.

For this reason, the questions investigated in this study concern only the economics of the generation of the new technologies and not their subsequent redistribution.

The results show that the data are consistent, both before and after controlling for the innovations' places within technological trajectories, with systematic differences in innovation across sectors as predicted by a broad interpretation of the linear hypothesis. At the same time, these preliminary results reemphasize warnings made in the literature against assuming a simple one-to-one relationship between basic and applied innovation, and provide clues for a more realistic albeit a more nuanced model of the innovation process to aid in considering policies for the different R&D sectors in biotechnology and agriculture.

This paper proceeds in Section 2 by reviewing the economic and business literature to develop hypotheses that integrate evolutionary theories about the micro-patterns of heterogeneous technological generation with organizational theories of comparative advantage in R&D. Section 3 then presents the patent data, Section 4 develops the econometric test, Section 5 presents the results, and Section 6 concluded with the major lessons learned.

2. Framework for analysis

2.1 The theory of micro patterns in innovation: technological trajectories

Micro patterns of innovation have long been implicated in empirical studies of the determinants of R&D output (reviewed in Cohen, 1995). In one of the earliest econometric studies to use patent data, Scherer (1965) found variation in patenting activities to be significantly related to the scientific or technical fields in which they occurred, with technological progress moving significantly more rapidly in some fields than in others. A variety of technological taxonomies have subsequently proven able to control for R&D efficiency measures of '*technological opportunity*' (Jaffe, 1986; Levin, et al., 1987).

Efforts to explain such field-specific discrepancies or patterns in the rates and characteristics of innovation have led to the concept of the natural technological trajectory. Rosenberg (1969; 1974) describes innovative efforts as focused on solving a finite set of closely or sequentially related problems which he terms *focusing devices* or *technological imperatives*—bottlenecks, weak spots, and clear targets for improvement—resulting in *compulsive sequences* of innovations over time. Somewhat more focused on final markets, Abernathy and Utterback (1978; Utterback, 1979) describe a *technology life cycle* with four phases: (1) the early experimental pre-paradigmatic phase, (2) the emergence of a *dominant design*, (3) the mature phase of refinement in which incremental innovations decrease costs and exploit economies of scale, and finally (4) the phase of decline and obsolescence, until the dominant design is replaced by radically new technologies and a new cycle begins again. Nelson and Winter's (1977) notion of *technological regimes* growing over time along *natural technological trajectories* contains may of the same elements discussed by Rosenberg and others and borrows, in addition, from the notion of R&D as a search mechanism (Evenson and Kislev, 1976) and induced innovation theory (Binswanger, 1974; Hayami and Ruttan, 1985). Concerned with the relative inflexibility built into induced innovation models by deterministic conceptions of decision making in the R&D process, Nelson and Winter seek in their heuristic natural trajectories model to balance the simultaneous influences of *demand pull* and *technology push* in an explanation of the patterns of R&D output. They observe that R&D strategies adjust to the incentives and constraints of changing demand and cost conditions faced by the commercialized outputs of R&D as well as the fact that initiation and success of a given R&D project is a function of the expected time, cost, and feasibility of the project, which in turn depend on the general state of science and the technological knowledge base of the researchers and engineers being employed.

Sahal (1981) and Dosi (1982; 1988) take the concept further and characterize the technological regime as the set of parameters of the meta-production function of Hayami and Ruttan, the set of potentially feasible yet costly technological capabilities traded-off by the technology user under physical or budget constraints. Dosi argues that this is equivalent to ascribing a set of hedonic attributes (Lancaster, 1966) to technologies, locating a particular set of coordinates in *technology characteristics space* around which individual innovations cluster to define a technological regime, either in the form of quantitative performance-cost characteristics as emphasized by Sahal or more cognitive or conceptual characteristics emphasized by Nelson and Winter.

These theories suggest that new innovations arise as results from different points along a spectrum of R&D modes. The R&D mode at one end of the spectrum tends to be of a more original and exploratory nature, testing the limits of the possible and probing the frontiers of known technology characteristic space. Most of the outcomes of such original exploratory research are dead-ends. Occasionally, however, one of these exploratory searches may happen upon a particularly promising problem-solving paradigm—in both the sense of creating new technological opportunities and in the sense of showing new ways to meet market demand—and may initiate the R&D mode closer to the other end of the spectrum creating follow-on innovations. The original work may, then, in hindsight come to be considered as having been a breakthrough discovery or a radical innovation.

As the new idea and the attendant technical information for a successful problem-solving paradigm diffuse (either directly or indirectly) to other investigators working in the same area, success can beget success. Competitors may notice the threat of a new approach in solving an old problem and attempt with new vigor to build upon or to work around the ideas of the initial innovators. The diffusion of the new paradigm continues with the making of numerous refinements and improvements clustered at those coordinates in hedonic technology characteristic space that were first pinpointed by the original breakthrough.

As this focused cluster of innovations accumulate over time, they form a *technological trajectory* along the time axis at that set of coordinates within technology characteristic space. The generation of successful and prominent technological trajectories continues to be driven both from the innovation supply side, by each new development in the trajectory and its associated cost reductions, and drawn from the demand side, with express customer demand manifest in the adoption both by midstream technology users and by final consumers of the products created with or embodying the new technology.

Innovations made early in the natural technological trajectory, under the first (exploratory) R&D mode can be expected to exhibit greater public goods characteristics compared innovations made later in the trajectory under the second (trajectory-building and -maturing) R&D mode. Technologies early in a trajectory are almost by definition more difficult to appropriate, as it is the very need to create more refined and profitable versions of the original good idea that drives the development of a technology forward along such a trajectory. Conversely, the later follow-on incremental innovations can be expected to exhibit greater private-goods characteristics, as they consist of focused applications intentionally designed to generate more appropriable returns in the marketplace. Moreover, when the exploratory innovations early in a trajectory were being made, there was much greater uncertainty as to whether or where research results might lead, and many in fact led nowhere. By contrast, those innovations being generated later in an established trajectory are by definition much more certain, yet they are also much more incremental and subject to greater inertial tendencies. However, a simplistic differentiation between characteristically public-good innovations and private-good innovations can and has been misleading.

At least since the watershed policy treatise of Vannevar Bush in 1945 these general distinctions of the steps in the innovation process have been conventionally labeled as *basic* versus *applied* research. This distinction is enshrined in the annual national R&D statistics reported by the National Science Foundation (NSF), which describes basic research as that primarily intended "to gain more comprehensive knowledge or understanding of the subject under study, without specific applications in mind" (National Science Board, 2000). Yet, as the trajectory notion emphasizes, whether R&D is basic or applied is not a simple black and white question; there are many shades of gray in between. There are also many feedback loops along a technological trajectory, and the convergence of different trajectories may give rise to valuable new technological applications. One might rather characterize the steps—the individual R&D projects—probabilistically, ranking their likelihood of producing both deeper understanding and more useful technologies, as simultaneously embracing both abstract issues of fundamental knowledge and targeting specific solutions to concrete problems (Stokes, 1997). Likewise, any piece of new knowledge or technology that is the output from an R&D project can simultaneously exhibit the telltale characteristics of both public and private economic goods[1]. Thus, a more nuanced reinterpretation of the linear model—the convention that basic science precedes and gives rise to applied technologies—might be that the earlier steps along the trajectory of an innovation process are more likely to exhibit public-good attributes (and therein to be more "basic"), and the later steps in the same trajectory are more likely to exhibit private-good attributes (to be more "applied").

Finally, it is important to note that the working definition of a natural technological trajectory may be drawn more narrowly or more broadly. The individual trajectories of several complementary component technologies can be aggregated together into a larger trajectory that describes the development of an entire technological system. Even the technological evolution of an entire industry may be considered an aggregate technological trajectory. The concept has been usefully applied at an intermediate level of analysis to examine the micro-patterns of new product innovations in recent industry studies of chemicals (Achilladelis and Antonakis, 2001) and telecommunications (Garrone, Mariotti and Sgobbi, 2002) among others.

---

[1] An argument made by Richard Nelson. See for example "What is 'Commercial' and what is 'Public' about

In the field of agricultural biotechnology, examined here in more detail, interesting examples of technological trajectories at an intermediate level of analysis include the following:

- the suite of techniques used for plant genetic transformation, a technology which enables most of the leading products of the industry;
- the Bt insect-resistance technology, found in products such as Bt corn and Bt cotton, which biologically protects crop plants against insect damage, replacing chemical pesticide sprays;
- herbicide resistance technology, found in products such as RoundUp Ready soybeans, which selectively allows crop plants to survive the application of a chemical herbicide that kills off all other plants, replacing soil tillage and other more costly methods of controlling for weeds; and
- the development of male-sterile parental lines, a technology that improves the efficiency of, and in some cases enables hybrid seed production.

---

technology, and what should be?" (Nelson, 1992)

2.2 Theories of organizational research capabilities

It should be noted that the preceding discussion contains no mention of the organizational or institutional nature of the agents—be they firms, individuals, universities, or governments—expending resources and R&D efforts to generate innovations. Two lines of discussion, one roughly traced to Schumpeter and the other roughly traced to Nelson and Arrow, tend to dominate in the economic literature exploring the differential capabilities of different kinds of organizations at generating innovations.

In the Schumpeterian tradition, discussion is largely focused on the private sector and the question of the relative advantages of established firms versus new entrants in innovation. Beginning with Scherer's (1965) treatment, empirical studies have considered the effect of a broad range of firm characteristics in addition to size and market power on innovation—often controlling for field effects usually defined in terms of technological opportunity—relating firm characteristics to (homogeneous) quantities rather than (heterogeneous) qualities of the innovative output. Suggestive exceptions include Henderson (1993), whose framework effectively relates empirically different qualities of innovation—radical versus incremental and technical versus organizational—to characteristics of the firm, demonstrating that larger incumbent firms are more likely to pursue incremental innovation and less likely to pursue (disruptive) radical innovations. Cohen and Klepper (1996) show firm size giving a comparative advantage in exploiting process innovations relative to product innovations. In this Shumpeterian tradition public sector research is typically regarded as merely an exogenous factor creating technological opportunity to be exploited by the private sector agents that populate the models.

The theoretical notions posed by Nelson (1959) and Arrow (1962) concern a different economic question—that of the socially optimal provision of innovation given the uncertainty, inappropriability, and public-goods nature of the knowledge created—and the discussions that follow their lead largely focus on the tradeoffs between public and private sector provision of R&D. Dasgupta and David (1994) examine how the different institutional structures and social communities of publicly supported '*open science*' and privately financed '*commercial technology*' influence the efficiency and output of these respective R&D systems. However it is Trajtenberg, Henderson, and Jaffe (1992) who propose empirical measures able to get at the more qualitative notions of "basicness" and appropriability of individual inventions and are thereby able to empirically show significant qualitative differences between the more basic outputs of university research versus the more applied results of corporate R&D.

2.3 Combining theories: organizational comparative advantages and a division of innovative labor within technological trajectories

The perhaps unrivalled breadth of involvement by the public sector in agricultural R&D, together with the new possibilities for empirical analysis with the kinds of patent measures introduced by the work of Trajtenberg, Henderson, and Jaffe, beg for a more comprehensive and integrated framework relating the qualities of innovating organizations to the qualities of their innovations. Researchers, regardless of the sector in which they work, can be considered to face a universal optimization problem: given the opportunities and incentives posed by their organizational environment as well as the budget and policy constraints they face, to maximize their own individual utility in pursuit of some combination of three fundamental goals: fame, fortune, and freedom. The specific incentives and constraints of the organizational environment include hiring and promotion practices, publication and peer review, salary and tenure (or seniority) ladders, as well as royalty-sharing and conflict-of-interest policies.

Given a choice of sectors, a researcher, given his skill set and his preferences over the different incentives offered, will self select into an organization with the system of incentives and constraints that he expects will allows him to pursue the kind of research that will maximize his individual utility. Once employed, a researcher makes specific choices of research projects, given that system of incentives and constraints. At the same time, the management of the organization constructs an organizational environment consisting of incentives attractive enough and constraints reasonable enough to engage talented researchers and induce them to be as innovative as possible in those kinds of research outputs that will maximize the benefits to the organization, its shareholders, or it constituents.

In open science, to echo Dasgupta and David, research employees are provide with incentives that have evolved to meet the university's or government laboratory's institutional set of objectives, taking into account the fact that faculty or research staff are all the time pursuing their own individual objectives of fame, fortune, and freedom. Government agencies and universities typically strive to allocate their limited budgets as to maintain as many quality programs in as many fields as possible. In light of this constraint, they prioritize original research contributions and give their research employees sufficient freedom and opportunities to earn (at lest some degree of) fame for their successes. Academic fame may also result in (at least some degree of) personal wealth, in terms of higher salaries, more lucrative opportunities for consulting or other outside pursuits. However, this arrangement with less ‘fortune’ and more ‘freedom’ in the open science institutional setting gives rise to the potential for conflicts of interests. The necessary and appropriate institutional response is embodied in conflict-of-interest policies to provide guidance and constraints for balancing these two objectives, assuring that pursuit of ‘fortune’ by some employees does not crowd out the collective ‘freedom’ of all others. Given this structure of incentives and constraints, individuals who are driven by their own fascination, who are most able to take advantage of the opportunities provided by academic freedom, who value the opportunities that it creates for achieving individual fame, and who are willing to accept the lower odds of fortune will self-select and accept academic positions.

In industrial R&D, the alignment of opportunities for the pursuit of freedom, fortune, and fame arises mainly from firms’ need to show short-term profitability and medium-term development of new lines of business. Since most industrial R&D efforts are quite targeted, firms tend to give their scientists less freedom than do universities. However, base salaries are typically higher and industry scientists do not have the same hassle of constantly pursuing grant money. Depending upon line of business and firm size, job security in industry, where worker turnover is higher, may be less stable than in academia, where the averagely successful professor can attain tenure after mid career. On balance, the fortune of researchers in industry is higher, if somewhat more insecure or uncertain.

The ultimate result is that different types of organizations become endowed with different kinds of research talent and differently optimized strategies for maximizing benefits. Combined with the heterogeneity of research opportunities, defined by where a given research project lies within the evolving trajectories of knowledge and technology, it follows that differently endowed sectors will specialize according to their comparative advantage in generating research results with different qualitative characteristics. Research project choices include whether to explore uncharted territory or to pursue work within an already established technological trajectory, whether to attempt uncertain original experiments or to make more certain incremental advancements, all the while factoring the probabilities of success and the expected payoffs in terms of fame, fortune, and freedom.

2.4 Development of hypotheses

Dasgupta and David's two sectors of *open science* and *commercial technology* effectively describe the two most broadly general alignments of incentives and constraints for R&D. However, to effectively summarize the major organizational arrangements observed in biotechnology, these need to be expanded by dividing commercial technology further into two sectors: which we might call *'technological entrepreneurship'* and *'corporate R&D'*, reflecting the distinction common in the Schumpeterian tradition between new entrants and incumbents. Then, among these resulting three R&D sectors, a division of innovative labor within an identified technological trajectory can be hypothesized, based on heterogeneity in the characteristics of the research output including the timing (or age) and pace of innovation, as well as the scope, value (or quality), generality, originality, and appropriability of the technology.

Age: While priority in discovery is important for researchers working under all three regimes, it is hypothesized that earlier inventions within a given technological trajectory are more likely to arise from researchers in universities and government labs. These researchers have greater incentives to do initial work in exploratory and unestablished areas, given the driving criteria they face for creativity and self-differentiation. Then, the middle phases of a trajectory are more likely to arise from entrepreneurial firms, as the common business model in biotechnology involves startups backed by venture capital to explore technically proven but still uncertain commercial opportunities. Finally, corporations are hypothesized to be more likely to innovate in the later phases of more established trajectories, refining and scaling up technologies for market. Thus, considering age alone, this hypothesis describes the classic *linear model* of R&D. (See this and following hypotheses summarized in Table 1.)

Scope: The scope or breadth of individual inventions is hypothesized to be widest in open science, less so in entrepreneurial biotechnology, and most narrow in corporate R&D. The rational for patenting university and startup technologies is to market them, and the broader the patent the better its licensing or development potential. Also, public sector or startup firms' budgets for filing patent applications are often more constrained and therefore fitting more "invention per patent" can help conserve resources. Corporate researchers and their corporations may have incentives aligned to the opposite effect, with more patents per invention making the researcher's bonus check larger and the corporation's patent portfolio appear larger and stronger.

Value: Typically the distribution of values of inventions is highly skew (Scherer, Harhoff and Kukies, 2000). The few high value successes tend to be paradigm-setting patents that dominate in large areas of follow-on innovation and product development. Such patents are often the unexpected results of exploratory research or research directed toward other questions. Since the breadth of sampling in open science is greater and because such researchers may be more on the lookout for new ideas and applications, it is expected that the probability of occurrence of top value inventions would be higher in universities. Yet, the stakes and uncertainties of investing in such research and the desire to capture the value of such top value inventions is so great that the entrepreneurial sector may in fact be more likely to actually pursue and patent such inventions. This might be called the *'value filter'* hypothesis: venture capitalists and biotech startups will bet their investments only on the cream of the crop, and the entrepreneurial biotechnology sector may show the most valuable patents, followed by the open science sector. The incentives and dynamics of the corporate R&D sector seem less likely to consistently generate top value inventions.

Generality: Technologies with greater *'generality'* are those that drive follow-on innovation among a wider diversity of technological trajectories. Using a measure of the diversity of technology fields from which a patent is cited, Trajtenberg, Henderson, and Jaffe (1992) find university patents to be somewhat more general than corporate patents. This makes sense if one considers that—on the one hand— the value of a more general invention is likely greater but—on the other hand—it is likely more difficult to appropriate. Given the diversity of research programs found in open science organizations, their interest in broad social impact of results, and their lower regard for the appropriation of returns, it is hypothesized that organizations in the open science sector should be more likely to generate such measurably 'general' inventions.

In addition, many among the current generation of startups in biotechnology are created around expertise in general technology platforms, such as micro-array or genetic sequencing technologies, and they essentially sell the services of that platform.

In contrast, corporations, which are more focused on final markets, are hypothesized to be least likely to innovate in general technologies.

Originality: The breadth of prior knowledge on which an invention draws is proposed by Trajtenberg, Henderson, and Jaffe (1992) to define its *'originality'*. The assumption is that, since all new ideas are influenced by existing knowledge, drawing on a broad base versus a narrow base indicates more original or synthetic thinking[2]. Since originality or breadth of thinking is a key criterion of success in open science it is hypothesized that original inventions are more likely to be observed coming from universities and the like. Since breadth of inspiration is not as important a criterion in the incentives of corporate R&D, original patents are least likely to come from that source. Startups are presumed to be somewhere in between.

Pace of Innovation: Within a technological trajectory the pace of innovation, measured as the average lag time between citing patents, will presumably be slower earlier on, as larger conceptual and technological feasibility issues are being worked out. The pace likely quickens as innovation in that trajectory becomes more routine and as competition intensifies to get products to market. University and government research is

[2] Arguably, an invention that draws on no prior knowledge at all is original. The definition based on breadth of influence conforms to the measure that is available to test this hypothesis. See Table 1.3.

thus less likely to be associated with patents that have short lag times, while it is not clear whether startups or corporations will have an advantage in fast paced innovation.

Appropriability: Defined as the degree to which a patent builds upon the existing technologies already owned by the same organization, this measure of the *'appropriability'* reflects a technology's dependence upon the internalized transfer of knowledge: an appropriable technology does not transmit readily through external spillovers. Conversely, the greater the degree of spillovers from a technology, the smaller the proportion of its value that is left to be appropriated by the inventing organization. The building of protective 'patent fences' by filing extensively around a valuable patent position is a commonly discussed strategy in intellectual asset management. It is thus hypothesized that both types of private sector R&D organization will strongly emphasize innovation with high appropriability, while researchers in open science will be much less concerned with appropriation or building upon their organization's own prior patents.

Table 1. Summary of hypotheses

Characteristic of invention:	**Sector of inventor:**		
	Open science	**Technological entrepreneurship**	**Corporate R&D**
Age or priority of the invention	+++	++	+
Scope of the invention's claims	++	++	+
Value of the invention	++	+++	+
Generality of application	+++	++	+
Originality of idea	+++	++	+
Pace or rate of innovation	+	++	++
Appropriability	+	+++	+++

+++ is most likely, ++ less likely, and + least likely sector to specialize in each characteristic.

3. The data

The data for this study were combined from two separate collections of U.S. utility patents encompassing all inventions made in the life sciences over the last 30 years with relevance to crop agriculture. The first data set, assembled in 1999 from MicroPatent data, contains 2477 U.S. patents granted from 1973 to 1998 (De Janvry, et al., 1999; Graff, Rausser and Small, 2003). The second data set, assembled by Aurigin Systems in late 2001, consists of 4303 U.S. patents granted between 1982 and 2001 (Graff, et al., 2003). The intersect of the two sets is 1677 patents, yielding a combined collection for this study of 5103 U.S. patents granted between 1973 and 2001. Patents in both of the original data sets were selected using complex iterative data base searches over patent classification numbers, technology terms, and patent citation links and were thoroughly screened by experts in the field of plant biology, in an exhaustive effort to include all patents with pertinent subject matter but to exclude any patents with non-pertinent subject matter[3].

3.1 Determination of technological trajectories

This chapter presents only a crude first pass at the problem of empirically identifying technological trajectories. To begin to operationalize such an empirical question, I borrow from the conceptual framework of quantitative phylogenetics (biological systematics) (Schuh, 2000; Swofford, et al., 1996). Just as two organisms or species are more likely to

[3] The data selection and screening processes are described in further detail in Appendix A.

be evolutionarily related if they display a greater degree of homology or similarity of anatomical, physiological, or genetic characteristics—and thus to be categorized into the same phylogeny—so analogously are two patents more likely to be evolutionarily related (in an economic sense) if they display a greater degree of homology of conceptual or technological characteristics[4]—and should thus be categorized into the same technological trajectory. The methods employed in this current study are as of yet unable to compare the homology between two patents on a continuous scale, such as would be obtained by comparing genetic sequences of organisms; however, several discrete indicators of homology between patents are exploited, much as older phylogenetic methods (developed before DNA could be readily sequenced) employ discrete data on the anatomical or physiological characteristics of organisms. In taking a first look at technological homology using discrete data, I have explored specific methodologies from bibliometrics (scientometrics or information science) including (1) a topical indexation, in the form of the International Patent Classifications (IPCs) system, (2) co-word cluster analysis of the technical language in the text of the patents (Callon, et al., 1983; Callon, Law and Rip, 1986; Noyons and Van Raan, 1998), (3) co-citation cluster analysis of the patterns of citations made to older patents (Small and Griffith, 1974; Small, 1973; Zitt and Bassecoulard, 1996), and (4) analysis and categorization of patents by an expert in the field. The first and second methods, IPC indices[5] and co-word cluster analysis[6], are

[4] The analogy here between a 'species' and a 'patented technology' holds, as well, given that each fundamentally requires and embodies a criterion of uniqueness.

[5] The IPC based clustering process is described in Appendix B.

[6] The co-word cluster analysis is described in Appendix C.

applied to the full data set for this study. The fourth method, an expert analysis and categorization[7], has been carried out on just the 1973-1998 data set.

3.2 Determination of inventing sector

The question of *what sector has generated the new invention* was determined for each patent in the collection by examining the organization to which it was assigned when issued by the patent office. The names of such "assignees-at-issue"[8] are recorded in the patent data. However, several issues complicate the usefulness of the names thus obtained. First, something as simple as the consistent identification of an individual organization is complicated in a data set of 5,000 documents by the fact that an assignee may be listed under different names or under different spellings (and misspellings) of those names on different patents. Second, different business units or subsidiaries of a single larger organization may each receive patents in the name of the business unit or subsidiary rather than in the name of the larger organization. And, third, a small fraction, about 6 percent, of the patents have more than one assignee, and some of those involve collaboration across different sectors.

The approach taken to solve the first complication was to clean the names of the assignees, uniformly giving all patents for each assignee the single most-common spelling. In response to the second challenge, all documents assigned to a smaller entity

[7] The expert analysis and categorization is described in Appendix D.

[8] Inventors are the original owners of intellectual property rights, but they usually assign the rights to their employer. In the case that a patent's inventor is independent there may not be any assignee, and the patent simply remains the

known to have been majority owned by a larger entity at the time the patent was filed were reassigned in the data set to the parent entity. In response to the third complication, co-assigned patents were simply attributed to the first assignee listed on the patent, since priority listing often indicates the lead institution in a collaborative relationship.

Each assignee was then identified as a (1) university, (2) government agency, (3) non-profit organization, (4) individual, (5) small entrepreneurial firm, or (6) large corporate firm. The most difficult differentiation was between the last two, both because some medium sized firms defy easy classification as either "entrepreneurial" or "corporate", and because several of the most active small biotech firms in the industry were acquired by the large corporations in the industry during the timeframe considered. Since the fundamental research question seeks to relate organizational comparative advantages to innovative outcomes, the rules of thumb used to determine between these two categories considered issues of size, age, the nature of ownership and financing (privately held versus publicly listed), and the publicly projected culture of the firm. In the cases of acquired firms, patents assigned in the name of a small firm were tabulated as "entrepreneurial" if filed before the date that the firm was acquired by its corporate parent. Applications made after that date were then considered "corporate".

Table 2. The six types of organizations inventing agricultural biotechnologies

Type of R&D Organization	Patent Count	Percent of Total
Universities	957	18.8%

property of the independent inventor. However, only 1.7 percent of the documents in this data set went to independent inventors. (See Table 1.1.)

Government laboratories, agencies	291	5.7%
Non-profit research centers, foundations	37	0.7%
Individual inventors	89	1.7%
Entrepreneurial firms: biotech startups, small private companies	916	18.0%
Corporate firms: large, diversified, publicly listed	2,813	55.1%
Total:	**5,103**	**100.0%**

3.3 Independent variables

Not all inventions are created equal. The bibliometric indicators introduced in Trajtenberg (1990) and in Trajtenberg, Henderson, and Jaffe (1992) as well as other variations on these indicators have been linked to important aspects of economic heterogeneity in the technologies underlying patents (Hall, Trajtenberg and Jaffe, 2000; Harhoff, et al., 1999; Lanjouw and Schankerman, 1999). Table 3 lists the set of indicators employed in this study, defines each briefly, and describes what economic quality of the invention is measured or indicated by the variable. Most of these are taken directly from the NBER Patent Citations Data File, and detailed definitions are available in the reference paper that accompanies the data file (Hall, Jaffe and Trajtenberg, 2001).

Summary statistics of the Aurigin citations data and the NBER indicators are shown in Table 4. Annual averages are charted in Figures 1 thru 4 for citations received, self-citation ratios, generality, and originality. Since the NBER Data File ends in 1999, these indicators are available only for the subset of 3210 patents issued by 1999. Additionally, the NBER counts the citations received by 1999. Since additional citations have accumulated by 2002, counting these could be exploited to reduce the truncation of this important variable. Thus, counts of “citations made” and “citations received” were made by Aurigin Systems in April 2002 and used for all of the patents in the combined data set.

Correcting for truncation of “citations received”: The problem of truncation in the citations received variable is clearly illustrated in Figure 1. Correction for this truncation is made by multiplying the count of citations received by lag-specific inflators derived from the distribution of U.S. patent citation lags as in Hall, Jaffe, and Trajtenberg (2000), asking in essence what percentage of the patent’s lifetime-expected citations (based on the distribution of citations received by all U.S. patents) it has received by 2002. The actual number of citations received is then multiplied by the inverse of that percentage to project the number of citations the patent is likely to receive.

Table 3. Independent variable definitions and sources

Patent Variable	Definition	Indicates	Data Source
Age	Calculated from the application date of the patent	***Priority*** of the invention	calculated
Number of claims	Count of how many separately numbered ("independent" and "dependent") clauses the observed patent lists in its "Claims" section.	***Size*** or ***scope*** of the invention (Lanjouw and Schankerman, 1999)	NBER
Number of citations made	Count of how many patents the observed patent cites as relevant prior art in its "References Cited" section.	***Quality*** of the invention (Lanjouw and Schankerman, 1999)	Aurigin
Number of citations received	Count of how many later patents have cited the observed patent as relevant prior art, as of April 2002.	***Value, importance*** (Trajtenberg, 1990), or ***quality*** (Lanjouw and Schankerman, 1999) of the invention	Aurigin
Weighted citations received	Same as "Number of citations received by 2002", simply adjusted for truncation. (See discussion on next page.)	***Value*** or ***importance*** of the invention (Hall et al, 2000; Hall et al, 2001)	calculated from Aurigin
3 or less weighted citations	Equals 1 if "Weighted citations received" is 3 or less; equals 0 otherwise	Whether it is a ***low value, unimportant,*** or ***low quality*** invention	calculated from Aurigin
50 or more weighted citations	Equals 1 if "Weighted citations received" is 50 or more; equals 0 otherwise	Whether it is a ***high value, important,*** or ***high quality*** invention	calculated from Aurigin
Generality index	Ranges between 0 if the subsequent patents citing the observed patent are concentrated in a single technology class and 1 if they are spread out across separate classes.	***Generality*** of the invention: the ***breadth of technological impact*** resulting from the invention (Trajtenberg et al, 1992, 1997; Hall et al, 2001)	NBER
Originality index	Ranges between 0 if the prior patents cited by the observed patent are concentrated in a single technology class and 1 if the cited patents are spread out.	***Originality*** of the invention: the ***breadth drawn upon*** by the invention (Trajtenberg et al, 1992, 1997; Hall et al, 2001)	NBER
Average backward citation lag	The average age of the patents cited by the observed patent at the time it was granted.	The ***pace or rate of innovation*** in the area of the invention (Trajtenberg et al, 1992, 1997; Hall et al, 2001)	NBER
Self-citation ratio	The percentage of patents cited that are assigned to the same assignee as the observed patent.	***Appropriability*** of the invention: transmission of the knowledge is easier internally than via external spillovers (Trajtenberg et al, 1992, 1997; Hall et al, 2001)	NBER

Table 4. Summary statistics

Patent Variable	type	Obs.	Mean	Std. Dev.	Min.	Max.
Age	years	5103	5.59	4.35	0.69	27.35
Number of claims	count	3210	14.08	12.19	1	125
Number of citations made	count	5103	3.25	6.33	0	125
Number of citations received	count	5103	3.56	10.03	0	156
Weighted citations received	continuous	3430	10.09	19.06	0	284.40
3 or less weighted citations	dummy	3430	0.49	0.50	0	1
50 or more weighted citations	dummy	3430	0.03	0.17	0	1
Generality index	(0,1)	1747	0.26	0.27	0	0.83
Originality index	(0,1)	2719	0.23	0.27	0	0.85
Average backward citation lag	years	2726	7.49	5.92	0	102
Self-citation ratio	ratio (0,1)	2624	0.20	0.35	0	1

Figure 1. Annual average patent citations received

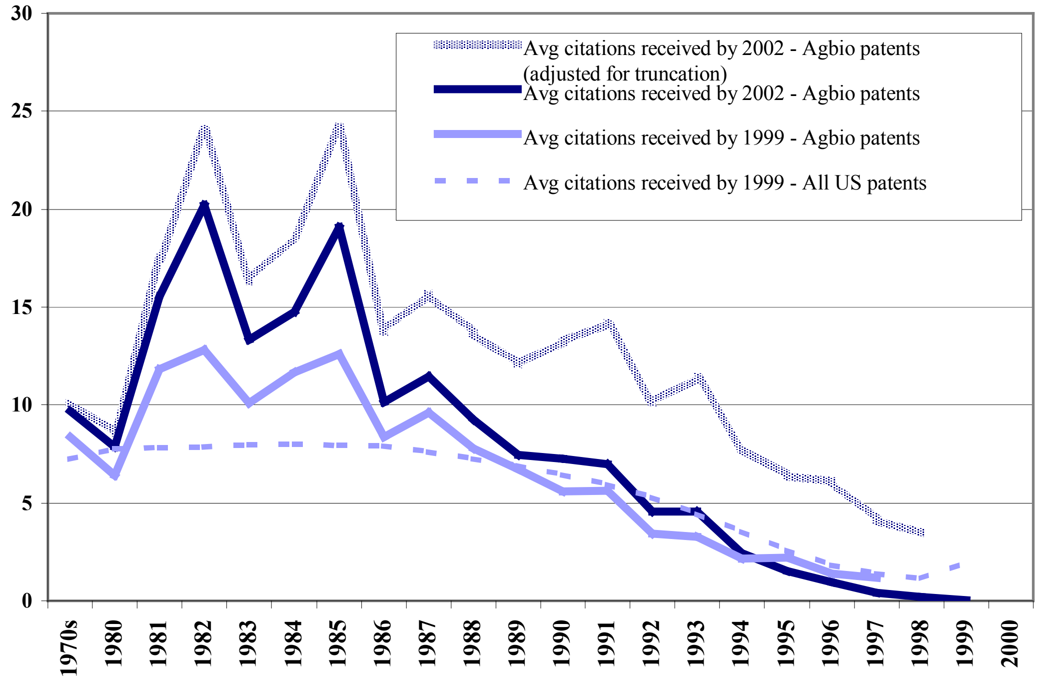


Figure 2. Annual average percentages of self-citations

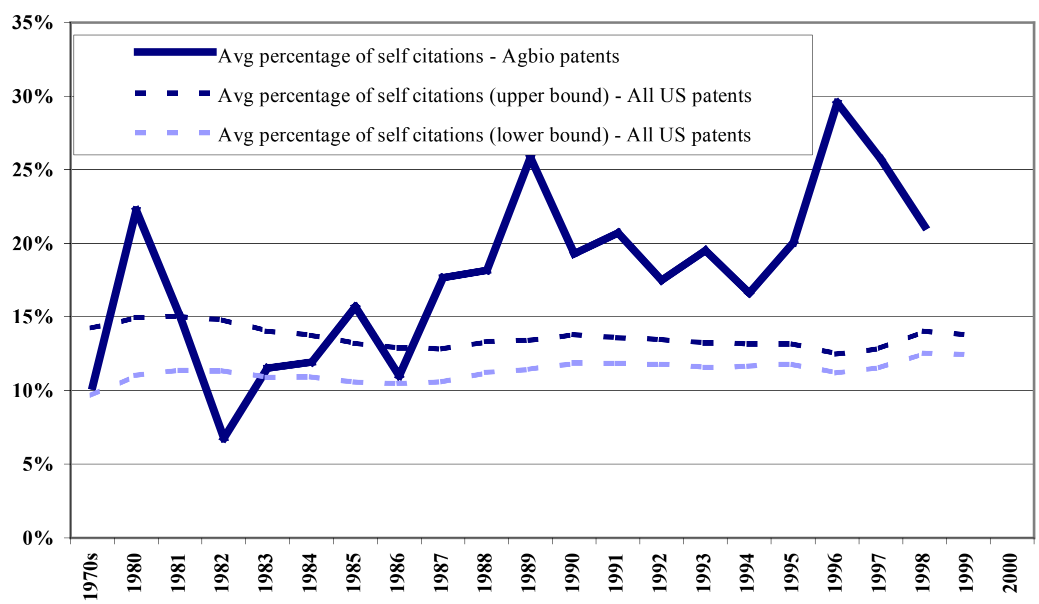


Figure 3. Annual average values of Generality index

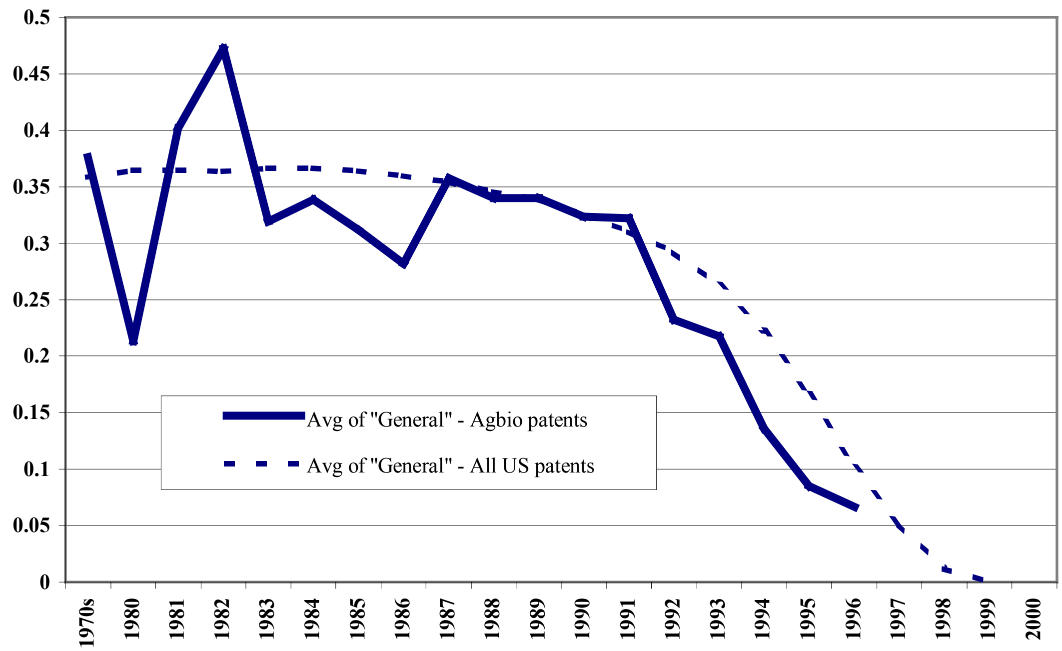
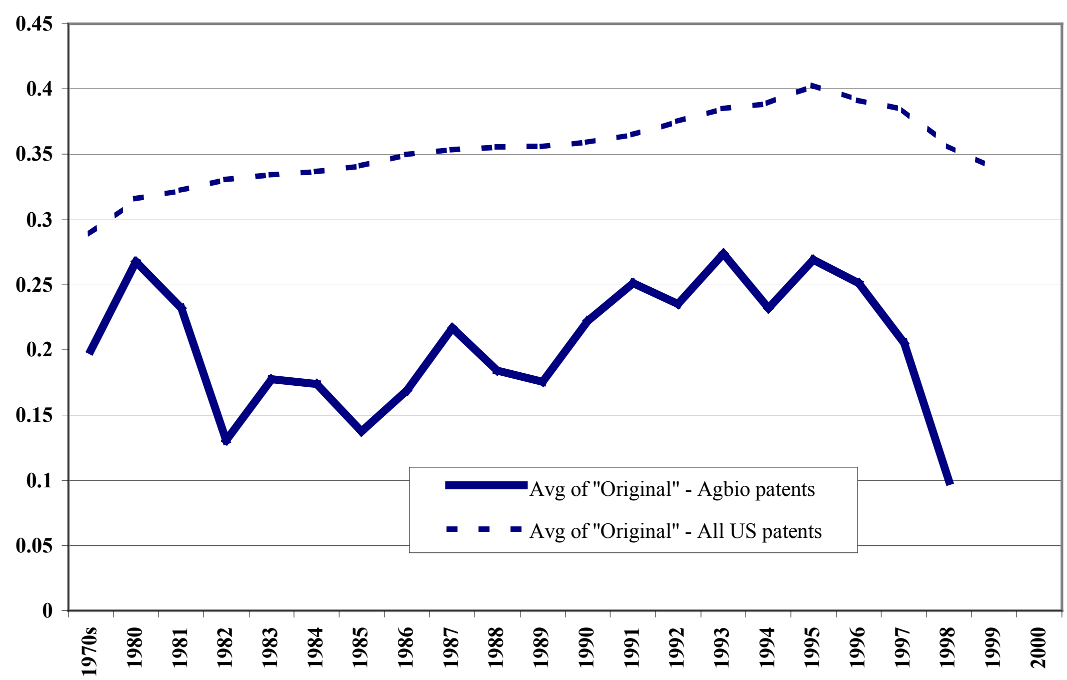


Figure 4. Annual average values of Originality index



4. The econometric model

4.1 The data generating process

There are J R&D sectors, $j = 1 \ldots J$, in the economy, distinct from one another in terms of the financing structures, prevailing cultures of open science versus proprietary technological R&D, and corresponding systems of incentives and constraints:

$$j = \begin{cases} 1 \text{ for universities, governments, and non-profit research laboratories,} \\ 2 \text{ for individuals, entrepreneurs, startup firms, and small businesses, and} \\ 3 \text{ for corporations.} \end{cases}$$

R&D is assumed to proceed within distinct research paradigms such that the resulting technologies are generated along naturally occurring trajectories, $k = 1 \ldots K$, each of which can be assumed to be ensconced within and thus captured by the classification of a technological sub-field or patent cluster. Not all technological trajectories are at the same point of maturity in their growth or evolution: some constitute new (and thus perhaps poorly defined) areas of research with little accumulated prior knowledge; others are mature areas with large stocks of existing knowledge already in place.

The underlying behavioral model of any researcher, in any sector, j, for generating research results and ultimately patents consists of several distinct steps:

1. A joint decision is made by a researcher and the research administrator or funding source in sector j to undertake a project in a specific research sub-field corresponding to an existing or emerging technological trajectory, k. It is presumed, although not observed, that the expected (joint) returns from this

research project exceed the expected returns from the next best deployment of the researcher's time and talents and the R&D organization's resources.

2. With a certain probability, a successful research result is produced that meets the standard criteria for patentability of being a novel, non-obvious, and useful.
3. Another joint decision is made by the researcher and their host organization as to whether the (novel, non-obvious, and useful) research result be patented, versus alternative strategies of being kept as a trade secret, being published in the public domain, etc. It is presumed, although again not observable, that the expected (joint) returns to taking a patent on this invention, subject to policy restrictions and transaction cost constraints, are greater than the expected returns to taking a patent on the next best invention. Thus, with a certain probability, or *patenting propensity*, a patent is applied for and granted on the research result.
4. The patent, n, the R&D sector, j, of the assignee, the technological trajectory, k, to which it contributes, and the qualitative attributes, $\mathbf{X}_n$, of the patent are all observed. Some of the X can be observed immediately after the patent issues, others only after some time has elapsed.

4.2 A polytomous statistical model

Econometric models of discrete random outcomes, such as multinomial probit and logit analysis, have been adapted and employed by economists to estimate latent variable models of choice behavior (McFadden, 1974; Ruud, 2000) in which each outcome is interpreted as the choice of an individual economic agent whose unobserved or 'latent' utility, construed as a random variable, is assumed to have been maximized by the observed choice, also a random variable, made relative to all other available options. McCullagh and Nelder (1983) argue that the statistical model employed to analyze joint sample distributions of polytomous data and the underlying behavioral model used to describe the unobserved latent variable are, however, indeed separate models, and in most cases the latent variable, while useful for the internal consistency of the behavioral hypothesis, is often unverifiable in practice. Given the data limitations and the behavioral complexity of the innovation phenomena addressed in this study, it is not possible to identify a single, behaviorally meaningful latent variable. Instead I simplify the complex decisions effected by the many unobservable parameters and latent behavioral variables at play in the data generating process into a single "black box" probability index that relates the qualitative characteristics of a patent with the probability that it is observed to arise from research conducted in a particular sector of the economy.

In essence, this exercise is the same as the classical statistical problem of drawing a randomly distributed sample, pulling *n* colored chips from *j* barrels. For each observed patent, *n*, in each technological trajectory, *k*, the probability index that the technology is found to be invented and patent by the *j*th organizational type is denoted by

$$\mathbf{y}^*_{nj} = \mathbf{X}_n\mathbf{B}_j + \boldsymbol{\varepsilon}_{nj},$$

where $\mathbf{X}_n$ is a vector of attributes of the n^{th} patent and the $\mathbf{B}$ are unknown coefficients. The ε_{nj} are the unobserved differences in the probability of that patent arising in the j^{th} type of R&D organization, resulting from unobserved features of the behavioral model including the institutional features of the organization, and are assumed to be i.i.d. random variables with a Weibull probability distribution.

When the j^{th} organizational type actually undertakes the research and receives the n^{th} patent, the observed outcome is described with the J dummy variables where

$$y_{nj} = \begin{cases} 1 \text{ if the } n^{\text{th}} \text{ patent is issued to the } j^{\text{th}} \text{ organizational type} \\ 0 \text{ otherwise.} \end{cases}$$

From the probability index equation the probability of the n^{th} patent coming from the j^{th} organizational type is

$$\begin{aligned} P_{nj} &= \Pr[y_{nj} = 1 \mid \mathbf{X}_n] \\ &= \Pr[y^*_{ni} \leq y^*_{nj}, \forall i \neq j \mid \mathbf{X}_n] \\ &= \Pr[\varepsilon_{ni} - \varepsilon_{nj} \leq (\mathbf{X}_{nj} - \mathbf{X}_{ni})`\mathbf{B}, \forall i \neq j \mid \mathbf{X}_n] \end{aligned}$$

Which is equivalent, given the assumptions made about the distribution of the εs, to

$$P_{nj} = P(y_{nj} = 1) = \frac{e^{X_n B_j}}{\sum_{i=1}^{J} e^{X_n B_i}}$$

This can be normalized and written as the multinomial logit:

$$P_{nj} = \frac{e^{X_{nj} B}}{1 + \sum_{i=1}^{J-1} e^{X_{ni} B}}, \qquad \forall i \neq j$$

where the values of the J different 'P's are conditional probabilities of a patent's occurrence in the J different sectors given the independent variables describing the patent's attributes.

Because they do not enter the probabilities linearly, the organizational coefficients on these patent attributes, the **B**s, cannot be interpreted directly. However, an interpretation is possible from the definition

$$\ln\left(\frac{\Pr_{nj}}{\Pr_{no}}\right) = X_n' B_{jo} \text{ where } o \neq j = 1, 2, 3$$

or more conveniently for interpretation

$$\frac{\Pr_{nj}}{\Pr_{no}} = \exp\left(X_n' B_{jo}\right)$$

which is the probability ratio (also known as the relative risk ratio) of a given type of patent arising from a research organization of type *j* relative to a research organization of type *o*. The *q* parameters in the vector **B** are the marginal effects of the *q*th regressor in X_n on the odds ratio. Finally, since the multinomial logit system is solved by maximum likelihood, testing hypotheses about coefficients follows standard methods based on the covariance matrix from the maximum likelihood estimation.

5. Analysis and results

5.1 Patterns of R&D output qualities in agricultural biotechnology

First, to explore the general significance of the patent indicators as predictors of R&D sector of invention and to test the hypotheses at the industry level, multinomial regressions were run on the entire data set. This treats the evolution of the entire industry as a single “trajectory” under the terms of the hypotheses. Results (Table 1-5) comparing probabilities of the public sector (Sector 1) and the corporate sector (Sector 2) show that original patents are more likely to come from public sector R&D, a greater number of claims or citations make a patent more likely to come from the public sector, low value patents, those with 3 or fewer weighted citations, are less likely from the public sector, and highly appropriable technologies, those patents with high self-citation ratios, are significantly less likely to come from public sector organizations. Comparing entrepreneurial (Sector 2) and corporate (Sector 3) probabilities, older patents are less likely to be from entrepreneurs, or, conversely, entrepreneurial sector patents tend to be younger than corporate patents. Again, the least valuable patents, those receiving 3 or fewer weighted citations, are significantly less likely from entrepreneurs than from corporations. Entrepreneurs' patents appear to be significantly higher on the generality index than corporate patents, a rather surprising result next to the age effect noted above, since given the truncation in generality (see Figure 3) a cohort of younger patents should score lower on the generality index. Also surprising, there is weak evidence that the patents of entrepreneurs are more self-citing than corporate patents.

Table 5. Combined regressions on the full dataset: coefficients displayed

Multinomial Regressions:	(1)	(2)	(3)
Sector 1- Universities, Governments, Non-Profits			
Age	0.006 (0.013)	-0.015 (0.009)*	-0.005 (0.009)
Number of claims	-0.015 (0.006)**	-0.012 (0.004)***	-0.008 (0.004)*
Number of citations made	-0.067 (0.022)***		-0.038 (0.012)****
Weighted citations received	-0.005 (0.006)		
3 or less weighted citations	-0.560 (0.194)***	-0.462 (0.097)***	-0.418 (0.097)**
50 or more weighted citations	-0.062 (0.538)	-0.505 (0.350)	-0.480 (0.351)
Generality index	0.081 (0.299)		
Originality index	0.653 (0.303)**	0.065 (0.181)	
Avg. backward citation lag	-0.002 (0.012)	-0.887 (0.166)***	
Self-citation ratio	-0.760 (0.234)***		-0.899 (0.165)***
constant			
Sector 2- Entrepreneurs, Startup Firms, Individual Inventors			
Age	-0.068 (0.014)***	-0.058 (0.010)***	-0.053 (0.009)***
Number of claims	0.005 (0.005)	0.007 (0.004)*	0.008 (0.004)**
Number of citations made	-0.004 (0.015)		-0.013 (0.010)
Weighted citations received	0.000 (0.004)		
3 or less weighted citations	-0.503 (0.198)**	-0.868 (0.103)***	-0.860 (0.104)***
50 or more weighted citations	0.436 (0.423)	0.456 (0.264)*	0.459 (0.264)*
Generality index	0.584 (0.287)**		
Originality index	0.034 (0.288)	-0.006 (0.183)	
Avg. backward citation lag	-0.008 (0.012)		
Self-citation ratio	0.332 (0.189)*	0.061 (0.142)	0.046 (0.142)
constant			
Pseudo R-squared	0.06	0.08	0.08
Observations	1227	2166	2173

Sector 3- Corporations is the comparison group

Regression coefficients displayed
Standard errors in parentheses
* is significant at 10%
** is significant at 5%
*** is significant at 1%

5.2 Patterns of R&D output qualities within technological trajectories

Three test cases on specific technological trajectories are presented. One tracks the evolution of the central "general purpose" technology of the industry: genetic transformation of plants. The other two plot the development of the leading genetic trait technologies thus far commercialized in crops: Bt insect resistance and herbicide tolerance. These are the most mature and the best-defined technological trajectories in the industry, and it stands to reason that the fit of the model is significantly better in each of these than in the full industry regressions. Results here should be considered to carry more weight, as now the patent attributes are being compared among highly homologous technologies, and cross trajectory effects are not muddying the results. In order to facilitate cross comparison of the effects of the independent variables, odds ratios rather than regression coefficients are displayed in the following regression tables.

Plant genetic transformation technologies: All three methods for determining technological trajectories proved useful in identifying this trajectory: IPC groups 17 and 19 were combined to bring together general genetic transformation with vectors and methods specialized for plant transformation (in Appendix B); the co-word mapping algorithm placed a preponderance of transformation technologies into map clusters 1, 1a, and 10 (in Appendix C); and plant transformation technologies are clearly identified in group 2 of the expert analysis technology classification (in Appendix D). Regressions were run separately for the patents identified under each of these three systems (Table 6).

Table 6. Regressions in the plant genetic transformation technological trajectory: odds ratio displayed

	By IPC group		By co-word map cluster		By expert analysis category	
Multinomial Regressions:	(1)	(2)	(1)	(2)	(1)	(2)
Sector 1- Universities, Governments, Non-Profits						
Age	0.881 (0.159)	0.920 (0.151)	1.029 (0.049)		0.888 (0.133)	0.963 (0.133)
Number of claims	1.012 (0.019)		0.992 (0.016)		0.994 (0.031)	
Number of citations made	0.854 (0.081)*		0.975 (0.043)		0.810 (0.193)	
Weighted citations received	0.997 (0.023)	1.002 (0.014)	1.008 (0.015)		0.948 (0.036)	0.962 (0.018)**
3 or less weighted citations	0.783 (0.577)		1.644 (0.844)		1.016 (1.135)	
50 or more weighted citations	1.902 (4.427)		0.831 (1.014)		6.948 (14.546)	
Generality index	0.392 (0.453)	0.426 (0.453)	1.148 (0.955)		68.669 (144.911)**	37.647 (68.425)**
Originality index	5.841 (5.906)*	2.069 (1.756)	2.027 (1.646)		52.522 (95.343)**	14.162 (20.090)**
Avg. backward citation lag	0.983 (0.079)		0.959 (0.038)		0.926 (0.122)	
Self-citation ratio	0.272 (0.323)	0.359 (0.411)	0.382 (0.282)		0.141 (0.227)	0.208 (0.331)
Sector 2- Entrepreneurs, Startup Firms, Individual Inventors						
Age	1.030 (0.161)	0.962 (0.141)	1.051 (0.053)		0.914 (0.145)	0.986 (0.127)
Number of claims	1.015 (0.06)		1.040 (0.014)		1.025 (0.026)	
Number of citations made	0.942 (0.067)		0.970 (0.046)***		0.853 (0.132)	
Weighted citations received	1.048 (0.022)**	1.020 (0.011)*	1.025 (0.014)*		1.046 (0.026)*	1.009 (0.011)
3 or less weighted citations	0.308 (0.278)		3.637 (1.909)**		2.076 (2.308)	
50 or more weighted citations	0.005 (0.017)		0.281 (0.376)		0.109 (0.176)	
Generality index	0.143 (0.147)*	0.355 (0.331)	0.851 (0.787)		0.886 (1.503)	0.968 (1.418)
Originality index	1.462 (1.319)	1.003 (0.773)	1.653 (1.454)		0.657 (1.055)	0.272 (0.355)
Avg. backward citation lag	1.070 (0.055)		0.987 (0.034)		0.908 (0.128)	
Self-citation ratio	5.215 (3.679)**	3.884 (2.517)**	0.621 (0.448)		2.122 (1.990)	3.128 (2.679)
Pseudo R^2	0.12	0.06	0.06		0.19	0.13
Observations	138	138	204		68	68

"**Sector 3- Corporations**" is the comparison group

Relative risk ratios displayed
Standard errors in parentheses
* is significant at 10%
** is significant at 5%
*** is significant at 1%

The most consistent results across these plant genetic transformation regressions show that more original transformation patents are more likely to come from the public sector, while higher value patents, in terms of weighted citations received, are more likely from the entrepreneurial sector. There is some indication that patents with a higher self-citation ratio are less likely from the public sector and more likely from the entrepreneurial sector, both of course compared to the corporate sector.

Bt insect resistance technology: The Bt trajectory was identified by bringing together IPC groups 5, 9, and 15 (in Appendix B); in the co-word analysis, map cluster 21 is almost entirely Bt (in Appendix C); and, the Bt subsets of groups 7 and 15 were used from the expert analysis categorization (in Appendix D). Two equations with variable choices that fit reasonably well were run separately for all three systems (Table 7).

A consistent result on the age variable—that older patents are more likely to have come from the public sector—lends support to the simple interpretation of the linear hypothesis. We are statistically unable to differentiate by age between entrepreneurial sector Bt patents and corporate sector Bt patents, although the point estimates indicate that older patents are more likely corporate.

Several other strong results are consistent with earlier observations. The more original Bt patents are more likely to have emerged from universities and government labs. Higher value patents, as indicated by the citations received variables, are more likely (and low value patents less likely) to have come from the entrepreneurial sector. Finally, the odds on the self-citations ratio shows highly self-citing Bt patents much more likely to be generated by startups than corporations.

Table 7. Regressions in the Bt technological trajectory: odds ratios displayed

	By IPC group		By co-word map cluster		By expert analysis category	
Multinomial Regressions:	(1)	(2)	(1)	(2)	(1)	(2)
Sector 1- Universities, Governments, Non-Profits						
Age	1.074 (0.039)**	1.053 (0.027)**	1.139 (0.048)***	1.103 (0.036)***	1.163 (0.091)*	1.149 (0.068)**
Number of claims	0.976 (0.012)**	0.982 (0.008)**	0.984 (0.02)	0.981 (0.015)	1.110 (0.07)	1.012 (0.035)
Number of citations made	0.935 (0.037)*		0.925 (0.047)		1.351 (0.16)**	
Weighted citations received	0.994 (0.01)		1.004 (0.015)		0.916 (0.053)	
3 or less weighted citations		0.872 (0.176)				
50 or more weighted citations		0.842 (0.783)		0.869 (0.262)		2.182 (1.515)
Generality index	0.572 (0.303)		0.514 (0.376)		0.605 (1.246)	
Originality index	3.441 (1.835)**	1.202 (0.437)	3.672 (2.796)*	1.029 (0.555)	0.869 (1.898)	2.771 (3.487)
Avg. backward citation lag		1.033 (0.018)*		1.074 (0.034)**		1.076 (0.072)
Self-citation ratio	1.746 (0.744)		5.741 (3.971)**		0.276 (0.629)	
Sector 2- Entrepreneurs, Startup Firms, Individual Inventors						
Age	0.958 (0.038)	0.981 (0.026)	0.980 (0.049)	0.985 (0.034)	0.94 (0.056)	0.977 (0.042)
Number of claims	0.983 (0.011)	0.989 (0.008)	0.983 (0.018)	0.995 (0.013)	1.03 (0.034)	1.005 (0.019)
Number of citations made	0.996 (0.03)		0.979 (0.042)		1.045 (0.075)	
Weighted citations received	1.019 (0.008)**		1.03 (0.014)**		0.991 (0.011)	
3 or less weighted citations		0.566 (0.113)***		0.535 (0.149)**		0.464 (0.17)**
50 or more weighted citations		5.317 (3.473)**				
Generality index	1.82 (0.972)		1.096 (0.814)		11.408 (11.372)**	
Originality index	1.533 (0.824)	1.511 (0.527)	1.554 (1.183)	1.447 (0.717)	0.64 (0.693)	0.988 (0.662)
Avg. backward citation lag		1.002 (0.019)		0.952 (0.033)		0.958 (0.041)
Self-citation ratio	8.037 (3.05)***		28.633 (18.755)***		22.138 (17.549)***	
Pseudo R^2	0.09	0.03	0.15	0.06	0.29	0.06
Observations	420	668	255	371	144	207

"**Sector 3- Corporations**" is the comparison group

Relative risk ratios displayed
Standard errors in parentheses
* is significant at 10%
** is significant at 5%
*** is significant at 1%

Table 8. Regressions in the herbicide tolerance technological trajectory: odds ratios displayed

	By IPC group		By co-word map cluster		By expert analysis category	
Multinomial Regressions:	(1)	(2)	(1)	(2)	(1)	(2)
Sector 1- Universities, Governments, Non-Profits						
Age	--	--	0.993 (0.124)	1.01 (0.071)	0.997 (0.116)	1.024 (0.052)
Number of claims	--	--	1.014 (0.034)	0.992 (0.024)	0.991 (0.021)	1.002 (0.015)
Number of citations made	--	--	0.893 (0.072)	0.952 (0.035)	0.883 (0.073)	0.884 (0.062)**
Weighted citations received	--	--	0.973 (0.026)		1.029 (0.015)*	
3 or less weighted citations	--	--		0.404 (0.222)*		0.421 (0.239)
50 or more weighted citations	--	--	1.796 (2.735)		0.345 (0.460)	
Generality index	--	--	1.502 (1.992)		1.83 (2.450)	
Originality index	--	--	0.182 (0.277)	0.097 (0.120)*	0.18 (0.197)	0.171 (0.154)***
Avg. backward citation lag	--	--	0.993 (0.124)	1.01 (0.071)	0.997 (0.116)	1.024 (0.052)
Self-citation ratio	--	--	1.014 (0.034)	0.992 (0.024)	0.991 (0.021)	1.002 (0.015)
Sector 2- Entrepreneurs, Startup Firms, Individual Inventors						
Age	--	--	1.412 (0.225)**	1.177 (0.100)**	1.157 (0.177)	1.034 (0.082)
Number of claims	--	--	1.024 (0.046)	1.048 (0.026)**	0.98 (0.034)	1.000 (0.023)
Number of citations made	--	--	0.767 (0.161)	0.876 (0.092)	0.734 (0.215)	0.581 (0.145)***
Weighted citations received	--	--	1.041 (0.028)		1.025 (0.019)	
3 or less weighted citations	--	--		1.268 (0.843)		0.354 (0.303)
50 or more weighted citations	--	--	0.025 (0.060)		0.272 (0.516)	
Generality index	--	--	20.939 (41.153)		2.127 (4.471)	
Originality index	--	--	0.446 (0.975)	0.754 (0.783)	0.149 (0.245)	0.132 (0.189)
Avg. backward citation lag	--	--	0.20 61	0.11 96	0.14 63	0.29 100
Self-citation ratio	--	--	1.412 (0.225)**	1.177 (0.100)**	1.157 (0.177)	1.034 (0.082)
Pseudo R^2	--	--	1.024	1.048	0.98	1.000
Observations	--	--	(0.046)	(0.026)**	(0.034)	(0.023)

Herbicide tolerance technology: No group in the IPC system was able to distinguish this technology, since several different kinds of molecular mechanisms are employed to achieve this trait. The co-word analysis, however, did identify and structure several smaller clusters of these herbicide resistance mechanisms within map group 19 (in Appendix C.) The expert analysis explicitly sought out such trait-specific technologies; herbicide resistance is group 8 (in Appendix D.)

The equations chosen and run for these two categorizations of herbicide resistance technology employed a more limited but more effective field of regressors (Table 1-8). There is some indication in the results that more valuable patents are more likely to come from public sector inventors (particularly in the first expert analysis category regression), while conversely low value patents, those with 3 or less citations, are less likely to come from public sector inventors. Once again, a higher self-citation ratio significantly indicates against a patent coming from the public sector. Across both categorization systems, older patents appear more likely to be from entrepreneurial inventors, although it is significant only in the regressions for the co-word based trajectory.

6. Discussion and conclusions

Based on these preliminary results, there certainly are systematic differences across R&D sectors in the attributes of the patents they file. Can these differences be interpreted as specialization in the qualitative parameters of knowledge production? Four significant conclusions emerging from these results begin to answer the question.

The first conclusion is one of weak support for the simple age-defined linear hypothesis. As defined, this hypothesis can only be tested within the trajectory regressions. There is strong indication that older Bt patents are more likely to be from public sector sources and weak indication that older herbicide resistance patents are more likely to be from entrepreneurial sources. In most cases, however, patent age does not appear consistent with the scenario of first public sector invention and then entrepreneurial invention before corporate invention. More tests on more technological trajectories are needed. Until then, acceptance or rejection of this formulation of the linear hypothesis cannot be definitive.

The second is the strong acceptance of the 'value filter' hypothesis. The entrepreneurial sector is clearly the most likely source of high value inventions (as well as the least likely source of low value inventions), both within the specific trajectories and in the industry at large. While the public sector is less likely than the corporate sector to produce low value patents, it is indistinguishable from the corporate sector in high value patents. Biotech companies and their venture capital financiers appear to have filtered out the highest value talent and succeeded in creating more of the leading technologies in agriculture.

The third conclusion is the strength of originality in university, government, and non-profit R&D. The results on the originality index for the public sector are the most significant and persistent throughout the study. The case for originality is further strengthened by the fact that the number of citations made is a negative predictor of public sector patents. Since the originality index is constructed from citations made, it has a slight positive correlation with that variable: if more citations are made, there are more of them to be spread out over more technology classes. As a result, the already high originality related with public sector patents is probably biased downward.

The fourth conclusion is the acceptance of the appropriability hypothesis, but with a twist. University, government, and non-profit organizations are, as hypothesized, significantly predicted by low self-citation ratios and thus by low-appropriability technologies. The twist comes in the result that startups are a much more likely source than corporations of high self-citation ratios and, by correlation, of high appropriability technologies. Moreover, corporations have on average much larger internal portfolios on which to build and from which to draw self-citations. Thus, we would expect self-citations to be biased higher for corporations. This resonates with the value filter hypothesis: biotech startups and entrepreneurs are looking for technologies that are not only more valuable but technologies upon which they are able to build and from which they are able to appropriate the value created.

The synthesis of these, as well as the other more nuanced results, suggests a world of agricultural R&D in which public sector researchers do the most original biotechnology work, and do it a bit earlier. Entrepreneurs make their entry in the private sector if they have a high value technology that promises to be highly appropriable, and they build upon it. The corporations undertake the most innovating, in terms of generating sheer numbers of patents, but they are of more moderate appropriability and tend to be on lower in value.

References

ABERNATHY, W. J., and J. M. UTTERBACK (1978), "Patterns of Industrial Innovation," *Technology Review*41-47.

ACHILLADELIS, B., and N. ANTONAKIS (2001), "The Dynamics of Technological Innovation: The Case of the Pharmaceutical Industry," *Research Policy*, 30, 4: 535-588.

ARROW, K. J. (1962), "Economic Welfare and the Allocation of Resources for Invention," National Bureau of Economic Research: Princeton University Press.

BINSWANGER, H. (1974), "A Microeconomic Approach to Induced Innovation," *Economic Journal*, 84940-958.

BUSH, V. (1945), "Science, the Endless Frontier," Washington D.C.: United States Government Printing Office.

CALLON, M., J. P. COURTIAL, W. A. TURNER, and S. BAUIN (1983), "From Translations to Problematic Networks: An Introduction to Co-Word Analysis," *Social Science Information*, 22191-235.

CALLON, M., J. LAW, and A. RIP (1986), *Mapping the Dynamics of Science and Technology*. London: MacMillan Press.

COHEN, W. M. (1995), "Empirical Studies of Innovative Activity," in *Handbook of the Economics of Innovation and Technological Change*, ed. by P. Stoneman. Cambridge: Blackwell, 182-264.

COHEN, W. M., and S. KLEPPER (1996), "Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D," *Review of Economics & Statistics*, 78, 2: 232-243.

DASGUPTA, P., and P. A. DAVID (1994), "Toward a New Economics of Science," *Research Policy*, 23, 5: 487-521.

DE JANVRY, A., G. GRAFF, E. SADOULET, and D. ZILBERMAN (1999), "Agriculture Biotechnology and Poverty: Can the Potential Be Made a Reality?," University of Rome at Tor Vergata.

DOSI, G. (1982), "Technological Paradigms and Technological Trajectories: A Suggested Interpretation of the Determinants and Directions of Technical Change," *Research Policy*, 11, 3: 147-162.

— (1988), "Sources, Procedures, and Microeconomic Effects of Innovation," *Journal of Economic Literature*, 26, 3: 1120-1171.

EVENSON, R. E., and Y. KISLEV (1976), "A Stochastic Model of Applied Research," *Journal of Political Economy*, 84, 2: 265-281.

GARRONE, P., S. MARIOTTI, and F. SGOBBI (2002), "Technological Innovation in Telecommunications: An Empirical Analysis of Specialization Paths," *Economics of Innovation and New Technologies*, 11, 1: 1-23.

GRAFF, G. D., S. E. CULLEN, K. J. BRADFORD, D. ZILBERMAN, and A. B. BENNETT (2003), "The Public-Private Structure of Intellectual Property Ownership in Agricultural Biotechnology," *Nature Biotechnology*, 21, 9: 989-995.

GRAFF, G. D., G. C. RAUSSER, and A. A. SMALL (2003), "Agricultural Biotechnology's Complementary Intellectual Assets," *Review of Economics & Statistics*, 85, 2: 349-363.

HALL, B. H., A. B. JAFFE, and M. TRAJTENBERG (2001), "The Nber Patent Citations Data File: Lessons, Insights, and Methodological Tools," Working Paper No. 8498. Cambridge: National Bureau of Economic Research (NBER).

HALL, B. H., M. TRAJTENBERG, and A. B. JAFFE (2000), "Market Value and Patent Citations : A First Look," Working Paper No. 7741. Cambridge, MA: National Bureau of Economic Research.

HARHOFF, D., F. NARIN, F. M. SCHERER, and K. VOPEL (1999), "Citation Frequency and the Value of Patented Inventions," *Review of Economics & Statistics*, 81, 3: 511-515.

HAYAMI, Y., and V. RUTTAN (1985), *Agricultural Development: An International Perspective*. Baltimore: Johns Hopkins University Press.

HENDERSON, R. (1993), "Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Alignment Equipment Industry," *Rand Journal of Economics*, 24, 2: 248-270.

JAFFE, A. B. (1986), "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value," *American Economic Review*, 76, 5: 984-1001.

LANCASTER, K. J. (1966), "A New Approach to Consumer Theory," *Journal of Political Economy*, 74, 2: 132-157.

LANJOUW, J. O., and M. SCHANKERMAN (1999), "The Quality of Ideas : Measuring Innovation with Multiple Indicators," Working Paper No. 7345. Cambridge, MA: National Bureau of Economic Research.

LEVIN, R. C., A. K. KLEVORICK, R. R. NELSON, S. G. WINTER, R. GILBERT, and Z. GRILICHES (1987), "Appropriating the Returns from Industrial Research and Development; Comments and Discussion," *Brookings Papers on Economic Activity*, 3: 783-831.

MCCULLAGH, P., and J. A. NELDER (1983), *Generalized Linear Models*. New York: Chapman and Hall.

MCFADDEN, D. (1974), "The Measurement of Urban Travel Demand," *Journal of Public Economics*, 3, 4: 303-328.

NATIONAL SCIENCE BOARD (2000), "Science and Engineering Indicators," NSB-00-1. Arlington, VA: National Science Foundation.

NELSON, R. R. (1992), "What Is 'Commercial' and What Is 'Public' About Technology, and What Should Be?," in *Technology and the Wealth of Nations*, ed. by N. Rosenberg, R. Landau, and D. Mowery. Stanford, CA: Stanford University Press.

NELSON, R. R., and UNITED STATES. AIR FORCE (1959), *The Economics of Parallel R and D Efforts: A Sequential-Decision Analysis*. Santa Monica, Calif.: Rand Corporation.

NELSON, R. R., and S. G. WINTER (1977), "In Search of a Useful Theory of Innovation," *Research Policy*, 636-76.

NOYONS, E. C. M., and A. F. J. VAN RAAN (1998), "Advanced Mapping of Science and Technology," *Scientometrics*, 41, 1-2: 61-67.

ROSENBERG, N. (1969), "The Direction of Technological Change: Inducement Mechanisms and Focusing Devices," *Economic Development and Cultural Change*, 181-24.

— (1974), "Science, Invention, and Economic Growth," *Economic Journal*, 8490-108.

RUUD, P. (2000), *An Introduction to Classical Econometric Theory*. New York: Oxford University Press.

SAHAL, D. (1981), *Patterns of Technological Innovation*. Reading, Mass.: Addison-Wesley Publishing Company

Advanced Book Program/World Science Division.

SCHERER, F. M. (1965), "Firm Size, Market Structure, Opportunity, and the Output of Patented Inventions," *American Economic Review*, 551097-1125.

SCHERER, F. M., D. HARHOFF, and J. KUKIES (2000), "Uncertainty and the Size Distribution of Rewards from Innovation," *Journal of Evolutionary Economics*, 10175-200.

SCHUH, R. T. (2000), *Biological Systematics*. Ithaca: Cornell University Press.

SMALL, H., and B. C. GRIFFITH (1974), "Structure of Scientific Literatures .1. Identifying and Graphing Specialties," *Science Studies*, 4, 1: 17-40.

SMALL, H. G. (1973), "Co-Citation in the Scientific Literature: A New Measure of the Relationship between Publications," *Journal of the American Society for Information Science*, 24265-269.

STOKES, D. E. (1997), *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington D.C.: Brookings Institution Press.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, and D. M. HILLIS (1996), "Phylogenetic Inference," in *Molecular Systematics*, ed. by D.M. Hillis, and C. Moritz. Sunderland, MA: Sinauer Associates, 407-514.

TRAJTENBERG, M. (1990), "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *Rand Journal of Economics*, 21, 1: 172-187.

TRAJTENBERG, M., R. HENDERSON, and A. JAFFE (1992), "Ivory Tower Versus Corporate Lab: An Empirical Study of Basic Research and Appropriability," Working Paper No. 4146. Cambridge, MA: National Bureau of Economic Research (NBER).

UTTERBACK, J. M. (1979), "The Dynamics of Product and Process Innovation in Industry," in *Technological Innovation for a Dynamic Economy*, ed. by C.T. Hill, and J.M. Utterback. New York: Pergamon Press.

ZITT, M., and E. BASSECOULARD (1996), "Reassessment of Co-Citation Methods for Science Indicators: Effect of Methods Improving Recall Rates," *Scientometrics*, 37, 2: 223-244.

APPENDIX A. Sample selection methods

The fundamental questions of sample selection are the *definition* of the characteristics of the relevant set and then *identification* of members that have those characteristics. This exercise has five parts: (1) define the technological field of interest as specifically as possible, (2) determine the scope of search, such as the national jurisdiction (i.e. the U.S., Europe, Japan, etc.) and the range of dates (i.e. 1982-2002), (3) create and run queries using technology search terms or patent classification index numbers, being careful to cast the net broadly enough not to reject patents relevant to the defined fields of interest (minimize Type I error), (4) filter the data set using query strings or by hand to eliminate excess patents that do not conform to the defined fields of interest (minimize Type II error), and (5) undergo further iterations of (3) and (4) informed by and building upon the results previously obtained.

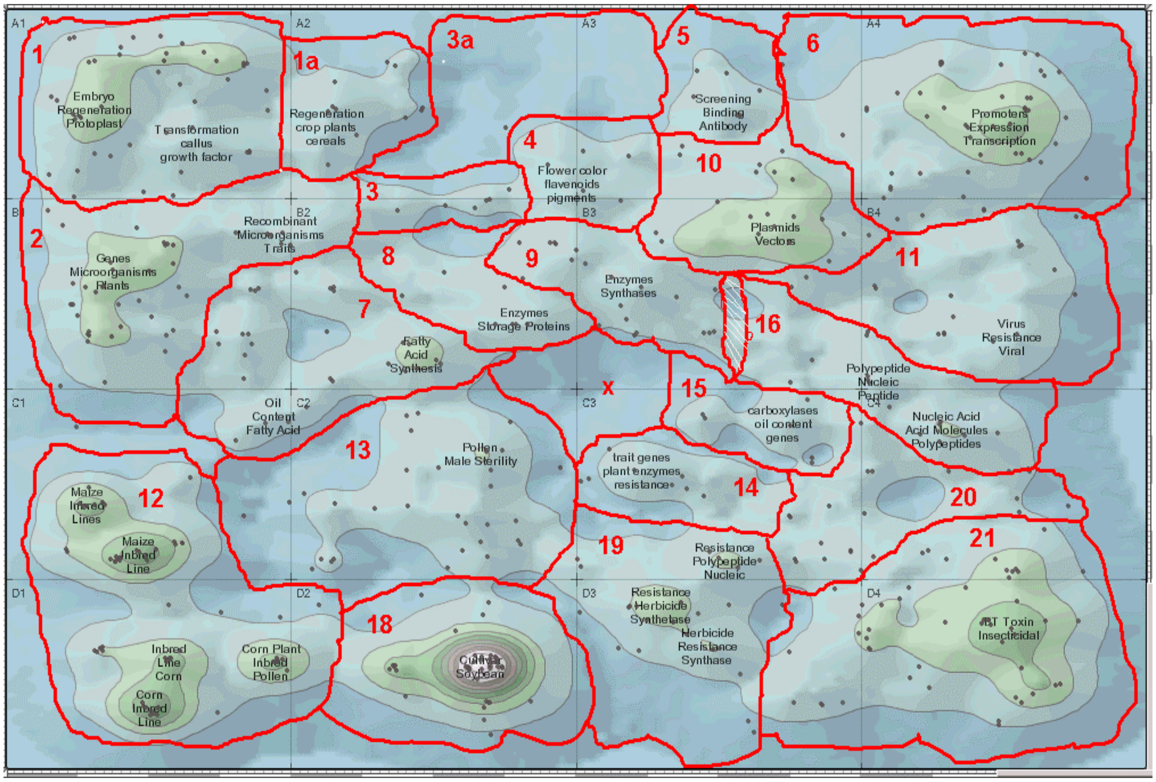
Steps:	Graff 1999 data set:	Aurigin 2001 data set:
Technology of interest:	Plant biotechnology, plant breeding, and biological control	Plant biotechnology
(1) Definition	Biologically based inventions in agriculture, including genetic engineering of plants, plant genes, plant varieties, biocontrol agents, plant breeding methods, etc.	Not explicitly specified
(2) Query method:	Based on US utility patents, years 1975-1998	Based on US, EU, JP, PCT utility and plant patents, years 1982 - 2001
(2a) search	Drew patents based with 1- one of a detailed list of (6-digit) US Classes, 2- specific technology keywords (from CABI Abstract index system) in title, abstract, or claims, and 3- English or Latin names of economically significant plants in title, abstract, or claims.	Drew all patents with one of the IPCs: A01H C12N C07K Drew all plant patents.
(2b) query based filtering	No query-based filtering.	***Level 1)*** *Enabling biotechnologies*: included C12N but excluded any A01H. Then include only those remaining patents that with general plant-related keywords in the title, abstract, or claims. Then exclude A61K, A23, and A21. Also exclude any food-crop-specific keywords. ***Level 2)*** *Generic plant technologies*: Include all A01H; exclude any food-crop-specific keywords. ***Level 3)*** *Food crop specific technologies*: Include only food-crop-specific keywords.

(2c) expert cleaning	Read through to eliminate non-agricultural and to categorize those kept (see 3 below).	Read through titles and abstracts to eliminate non-agricultural patents.
(2d) iterations	Examined citing "neighbors", those directly citing or cited by the selected set of patents. Then applied the cleaning step (2c) as above.	No iteration planned.
(3) Classification	Each patent assigned up to three categories (See Appendix D.)	A co-word based mapping algorithm was applied to cluster the similar patents in the data set together. (See Appendix C.)
The resulting data set	3092 US utility patents total; 2477 of these fell into the major categories of agricultural biotechnology.	4319 US utility patents total; 4303 of primary issue (i.e. not RE). (Also obtained 2911 European patents, 3685 Japanese patents, and 3479 PCT filings, not used in this study.)

APPENDIX B. International Patent Classification (IPC) groups identifying major technological trajectories in the agricultural life sciences

IPC (7th Edition) group definitions	IPC indices included	Patents
Horticulture; Cultivation	A01B 79/00, A01C **/**, A01G **/**	145
Plant breeding and hybridization	A01H 01/**	1123
Plant reproduction by tissue culture	A01H 04/**	749
Plant germplasm	A01H 05/00, A01H 05/02, A01H 05/03, A01H 05/04, A01H 05/06, A01H 07/00, A01H 05/08, A01H 05/10, A01H 05/12	1328
Biocides (*selected for biologically-based*)	A01N 63/00, A01N 63/02, A01N 63/04, A01N 65/00	340
Medicinal preparations (*plant-based)*	A61K **/**	149
Nucleic acids (genes and gene fragments)	C07H 15/00, C07H 15/12, C07H 17/00, C07H 19/00, C07H 21/00, C07H 21/02, C07H 21/04	332
Peptides	C07K **/**	184
Micro-organisms	C12N 01/**, C12N 01/15, C12N 01/19, C12N 01/21, C12N 03/00, C12R 01/**	770
(Plant) Cell lines and Tissues	C12N 05/00, C12N 05/02, C12N 05/04, C12N 05/10, C12N 05/12, C12N 05/14, C12N 15/02, C12N 15/03, C12N 15/04, C12N 15/05	416
Viruses (including genes encoding viral proteins)	C12N 07/**; C12N 15/33, C12N 15/34, C12N 15/35, C12N 15/36, C12N 15/37, C12N 15/38, C12N 15/39, C12N 15/40, C12N 15/42, C12N 15/51 (*highly dispersed across subclasses*)	70
Enzymes *(selected for plant-related)*	C12N 09/**, C12N 11/**	613
General genetic engineering	C12N 15/00, C12N 15/09, C12N 15/10; C12N 15/11	798
Genes encoding plant proteins	C12N 15/29	584
Genes encoding microbial proteins (e.g.Bt)	C12N 15/31, C12N 15/32	210
Genes encoding enzymes or proenzymes	C12N 15/52, C12N 15/53, C12N 15/54, C12N 15/55, C12N 15/56, C12N 15/57, C12N 15/60, C12N 15/61, C12N 15/62 (*highly dispersed across subclasses*)	291
Introduction of foreign genetic material using vectors	C12N 15/63, C12N 15/64, C12N 15/65, C12N 15/66, C12N 15/67, C12N 15/68, C12N 15/69, C12N 15/87	238
...**Vectors or expression systems adapted especially for microbes**	C12N 15/70, C12N 15/72, C12N 15/73, C12N 15/74, C12N 15/75, C12N 15/76, C12N 15/77, C12N 15/78, C12N 15/79, C12N 15/80, C12N 15/81	176
...**Vectors or expression systems adapted especially for plants**	C12N 15/82, C12N 15/83, C12N 15/84	735
Preparation of compounds or compositions by using micro-organisms or enzymes	C12P 01/**, C12P 07/**, C12P 09/**, C12P 13/**, C12P 17/**, C12P 19/**, C12P 21/00, C12P 21/02, C12P 21/04, C12P 21/06	449
Biological measuring or testing	C12Q 01/**, G01N 33/**	183

APPENDIX C. ***ThemeScape*** **map of co-word clusters from the combined dataset identifying major technological trajectories in agricultural biotechnology**



APPENDIX D. Classification of patents into technological trajectories based on expert analysis of the data set

Group number	Category definition	Examples	Number of patents
1	**Markers**: ▪ Selectable DNA/RNA markers ▪ Selectable/identifiable proteins ▪ Selectable/ identifiable phenotypes associated with genes of interest	marker sequences, applications, reporter genes, GUS, antibiotic resistance, antibiotics, leaf patterns, leaf size, colors, sprouting time	60
2	**Plant-specific transformation vectors and systems:** ▪ Agrobacterium, ti-plasmid ▪ Electroporation ▪ Biolistics, microprojectile ▪ Vectors, chimeric cassettes for insertion and expression ▪ Virus-based transformation systems ▪ Methods specific to a crop or taxonomic group ▪ Novel transformation methods ▪ Efficiency improvements ▪ Site-specific integration ▪ Plastid (chloroplast) or cytoplasmic integration ▪ Plant mutagenesis		216
3	**Bacteria-specific transformation vectors and systems**		55
4	**Promoters and regulation of gene expression**: ▪ Promotors; Expression; Amplification ▪ Transcription enhancement/ regulation/ suppression ▪ Sequence editing to modify expression ▪ Antisense/sense suppression ▪ Exogenous effects on gene function ▪ Inducible promoters (chemically/ environmentally induced, response to damage or infection) ▪ Tissue-specific expression/ suppression/ promoters; Developmental-stage-specific expression/ suppression/ promotors		284
5	**Plant cell, tissue, and embryo manipulation**: ▪ Cell, protoplast, and callus culture ▪ Somatic embryogenesis ▪ Organogenesis ▪ Plant regeneration from cell, protoplast, or callus ▪ Micropropagation, cloning ▪ Protoplast fusion ▪ In vitro selection, somaclonal and gametoclonal variation ▪ Microspore and macrospore plant culture ▪ In vitro sexual reproduction	somatic hybridization, cytoplasm transfer, selection, screening, varied culture conditions, high frequency embryogenesis, anther culture	237
6	**Plant disease/pathogen resistance**: ▪ virus resistance ▪ microbe resistance ▪ nematode resistance ▪ fungus/mold/mildew resistance ▪ hypersensitive response	Plant defenses, chitinase and glucanase expression	161
7	**Plant insect resistance**: ▪ Bt ▪ non-Bt	Includes all Bt genes & gene sequences	205

8	**Plant herbicide tolerance**: ▪ Bromoxynil tolerance ▪ Glyphosate (aka RoundUp®) tolerance ▪ Imidazolinone tolerance ▪ Phosphinothricin (Gluphosinate aka Liberty®) tolerance ▪ Sulfonamide tolerance ▪ Sulfonylurea tolerance ▪ Triazolinone tolerance ▪ Cyclohexanedione and/or aryloxyphenoxypropanioc acid tolerance ▪ Phenmedipham tolerance ▪ Aryloxyphenoxyalkanecarboxylic acid, imazethapyr, imazaquin, primisulfuron, nicosulfuron, sulfometuron, imazapyr, imazameth, imazamox, 3,5-dihalo-4-hydroxy-benzonitrile tolerance	Nitrilase from Klebsiella ozaenae, 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), Polypeptide into choroplasts, 5-enolpyruvyl-3-phosphoshikimate (EPSP) synthases, Acetolactate synthase, Phosphinothricin(PTC)-resistance gene, PAT (phosphinothricin acetyl transferase) from Streptomyces hygroscopicus, Acetolactate synthase, Aacetyl coenzyme A carboxylase (ACCase) carbamate hydrolase of Arthrobacter oxidans	119
9	**Plant physical/ agronomic performance traits**: ▪ nitrogen fixation ▪ photosynthesis ▪ nutrient/resource availability/ utilization/ apportionment ▪ plant morphology/strucural modifications/ organ modifications ▪ altered developmental pathways, life-cycle timing ▪ drought tolerance ▪ salt tolerance ▪ extreme temperature tolerance ▪ toxic metals tolerance ▪ Ph tolerance	seedlessness, short stem, altered flower, leafing patterns	41
10	**Control of plant reproduction** ▪ male sterility ▪ female sterility ▪ seed sterility ▪ apomixes ▪ self incompatability	creating hybrids, control of copying genetics, barnase from Bacillus amyloliquefaciens, GURTs, clones produced via seeds, diplosporous apomictic reproduction, diploid parthenogenesis, Tripsacum hybrids	95
11	**Primary nutrients quality and content enhancements**: ▪ amino acid, protein profile altered ▪ fatty acid, oil profile altered ▪ sugars, starch, carbohydrate profile altered	high lysine novel wheat glutenin, alpha-amylase expression, glycogen synthesis	185
12	**Other plant quality enhancements**: ▪ shelf life, ripening altered ▪ ornamental appearance altered ▪ fiber structure altered ▪ structural chemistry of plant altered ▪ Ph in plant cells altered ▪ altered solids or other components for improved processing ▪ reduced levels of harmful natural compounds in plants ▪ flavor compounds added, removed, suppressed, or altered ▪ vitamin or other micronutrient levels altered	ethelyne, fruit polygalacturonase levels suppressed cotton, flax fibers cellulose, lignin, wood chemistry high acidity tomatoes reduced nitrate levels in leaves sweetener proteins high carotene	142
13	**In-plant production of bio-molecules, compounds** ▪ production of industrial molecules, enzymes ▪ production of pharmaceutical molecules ▪ monoclonal antibodies ▪ production of vaccines ▪ production of nutritional molecules, vitamins ▪ production of flavorings, sweeteners ▪ production of herbicides/ plant toxins	"edible vaccines"	95
14	**Bio-based anti-pathogenic and plant disease treatment cultures or compounds:** ▪ Bio-based anti-viral plant treatments and compounds ▪ Bio-antibiotic plant treatments and compounds ▪ Bio-nematocidal compounds ▪ Bio-funcigidal compounds (+mold & mildew sprays)		171

15	**Bio-based insecticidal cultures or compounds**: ▪ Bt ▪ non-Bt	includes all Bt proteins & protein sequences (when not specifically claimed expressed in plants)	352
16	**Bio-based herbicidal cultures or compounds**	Xanthomonas campestris, Drechslera spp., Bipolaris sorghicola	61
17	**Hybrid plants**: ▪ parental lines ▪ inbred lines	primarily corn/maize	282
18	**Sexually reproducing plants**	primarily soybean	109
19	**Breeding and hybridization methods**: Breeding and selection Hybridization procedures and mechanisms	De-tassling, cross patterns	96